

Using Nvivo Qualitative Research

E Altmann 2013

School of Social Sciences, University of Tasmania

Altmann, E, "Using Nvivo Qualitative Research", *Social Research Methods*, Oxford University Press, Maggie Walter (ed), online, pp. online. ISBN 9780195520170 (2013) [Other Book Chapter]

Qualitative researchers work across a range of sectors to analyse structured, unstructured or semi-structured data such as interviews, surveys, field notes, web pages and journal articles, film and print media. They are interested in evaluating, interpreting and explaining social phenomena. While these activities can be undertaken without computer software, a growing number of qualitative researchers rely on a range of software packages to help manage and organise their data. Computer data analysis software for qualitative research became widely available to the public from the late 1990s (Rademaker, Grace & Curda 2012).

This chapter considers how computer assisted data analysis is able to assist qualitative researchers. It does this by focusing on one particular qualitative software package – NVivo. The chapter takes the reader through a stepped process that is similar to many other software packages, introducing terminology that qualitative researchers need to become familiar with. The strengths and weaknesses of NVivo are outlined, many of which relate to all computer assisted qualitative software packages. The chapter finishes with set exercises for the student seeking to learn more about the NVivo program.

How do computer programs assist qualitative research?

Computer assisted qualitative data analysis (CAQDA) are most useful in situations where large amounts of data transcripts need to be analysed. For example, where there are more than 25 interview transcripts to analyse, CAQDA is definitely useful. If you have less than 15 transcripts and are unfamiliar with the software package, then CAQDA will not be as useful. However rather than concentrating on the number of transcripts, the focus should rather be on the depth of information being investigated. A detailed interview with just one or two people over several days may be just as worthy of computer assisted qualitative data analysis. The decision to use computer assisted analysis will also be determined by the type of data being analysed. For

example, qualitative researchers may also have photographs, newspaper clippings, memos and reflective pieces about their interviews or participants that may also be used as part of the analysis. Thus the usefulness of a computer program is linked to the total amount of data rather than the number of interviews. An experienced researcher will be able to analyse the data more quickly without CAQDA assistance where there is a small amount of data.

CAQDA programs do not undertake the analysis as such. Rather, computer programs assist researchers whose work falls within two main areas. The first type generates theory by helping the researcher see relational links in the data. These programs support the development of new concepts and propositions. The second type assists in thematic analysis by providing code and retrieve functions. In qualitative data analysis, *coding* is the marking of data segments with categories, words or symbols that allow the grouping of similar categories of data. *Retrieval* is the process of accessing the data once it has been coded. Code and retrieve functions are an integral part of qualitative analysis. Coding provides an important way of organising material. The important point here is that what Geertz (1973) calls *rich data* is not dissipated in the process. Rather the data is taken apart, considered and then regrouped to show themes and relationships. Coding is good for identifying similarities in experience between participants.

Code and retrieve functions need to be undertaken in conjunction with rigorous *code book* development and the application of that codebook to text passages or data. Thus choosing the right computer program for the job is essential. This chapter uses *NVivo* to illustrate code and retrieve functions associated with CAQDA programs. It has been chosen because it is a relatively simple program to use. To do this you need to understand the capacity of the computer program and its strengths and weaknesses. You also need to understand your own capacity to manipulate the data within that program. Training in specific programs helps with both these aspects. Additionally you need an understanding of the data. As detailed in Chapter 14, you need to immerse yourself in the data. This can be done by reading and re-reading the transcripts to get a sense of what they contain and what might be important.

Focusing on NVivo Software

To assist students to understand how CAQDA can be of assistance to researchers, this chapter focuses on a computer package called NVivo. I have also used a dummy set of data to provide concrete examples. It is beyond the scope of this chapter to fully describe the program or provide tutorials in that program. However by summarising the most significant steps, a better understanding of how these programs are used is gained. NVivo Version 10 was released in June 2012 and provides the researcher with *theory generation* functionality and *code and retrieve* functionality. For McDonagh (2012), Nvivo allows the effective entering, coding and sorting of data. However Willis (2002) identifies NVivo as theory generating software. Newer versions of NVivo are both.

The first step is to setup the project in NVivo, naming and saving it. NVivo offers a choice of beginning a new project or opening an existing project. The number of steps that follow are project initialisation will in part be limited by whether you are using the CAQDA for generating propositions and theory, for code and retrieve functions or both.

Process steps:

- Prepare the documents for use and uploading external files

- Identifying themes

- Create nodes and beginning the coding process

- Explore themes and concepts

- Develop node system to illustrate relationships

- Reporting and write up.

Step 1 - Preparing documents for use and uploading external files

The decision to use CAQDA needs to be made at or prior to the time of transcribing the qualitative material. This is because there are implications for the way in which you format and save the data for importing into NVivo. NVivo allows you to create the sources in the program or import from other programs (Word for example). NVivo accepts a wide range of file formats such as audio files, videos, digital photos, Word, PDF, spread sheets, rich text, plain text and web and social media data.

Through the import and export functions NVivo also supports data exchange with applications like Microsoft Excel, Microsoft Word, IBM SPSS Statistics, EndNote, Microsoft OneNote, SurveyMonkey and Evernote, media files, web pages. One of the benefits of NVivo then, is that most relevant data formats can be stored within the one project file.

Qualitative researchers undertaking interviews often record and then transcribe the recorded interviews into text. NVivo provides several options here. You can for example upload *Word* and *wvm* files directly into NVivo. Some researchers may choose to code directly from the *wave* or *movie* files. There is a danger in approaching the data in this way as immersion in the data is a requisite of qualitative research. If the coding proceeds before the researcher becomes fully conversant with all the transcripts, some findings may be missed or data misinterpreted. Relational links may be lost.

To create an internal source in NVivo choose the *Navigation View*.

Click the *Sources* folder (or subfolder) under it.

Click on the *Create* tab,

Then click on the *Documents* tab.

The document properties *Dialog box* will open.

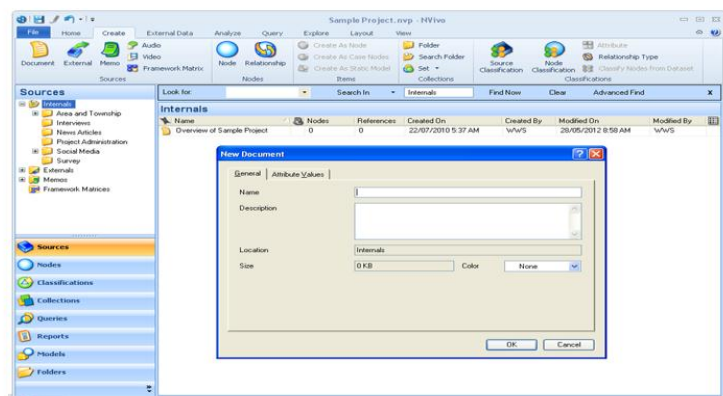


Figure 1 – Creating an internal source

You can directly enter text into this box for theory generation or coding and retrieval options. This option may also be useful for recording memos and field notes relevant to interviews. Memos provide context for the data and assist in the development of theoretical insights. By reflecting on the data researchers are better able to make connections between what was said or filmed and wider socioeconomic or political contexts.

Step 2 - Identifying themes

Once the data has been uploaded, a thematic search can be conducted. In Figure 2 the results of a thematic search on the uploaded data documents is shown.

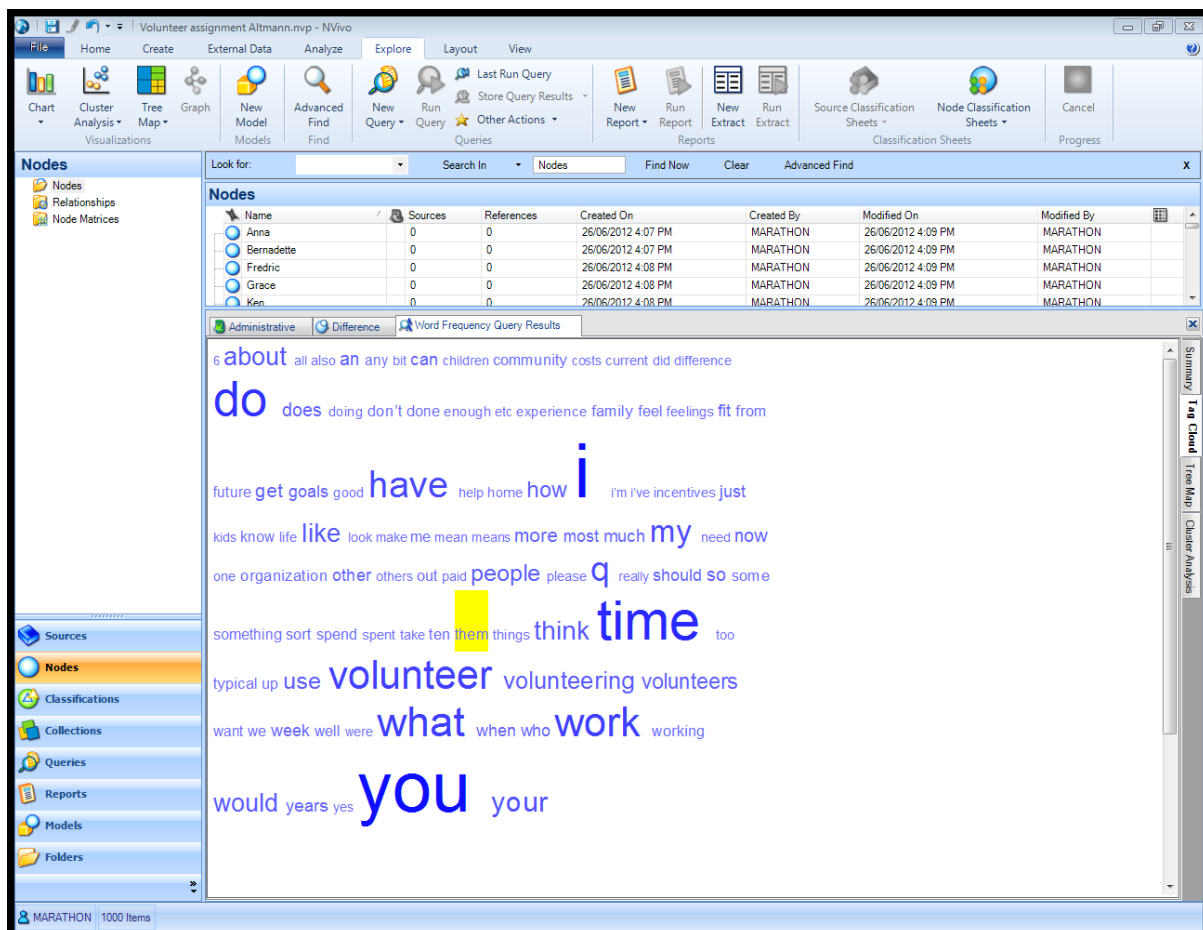


Figure 2 - A word tree within NVivo

NVivo has automatically counted the number of times a particular word has been used. NVivo calls this a *word tree*. The *word tree* assists in identification of key themes. The screen shot of this *word tree* indicates that the words *volunteer*, *time*

and *work* are used repetitively along with *I* and *you*. This is useful where large amounts of data have been uploaded and the researcher has little idea of what they are likely to find. The interviews may have been undertaken by someone else, or large volumes of web pages or news articles may have been uploaded. By clicking on the individual words, each reference can be scanned to briefly note the context in which the words were used by the participant. Where the coding of themes emerge from the data as in this case, it is called *inductive coding*.

Inductive coding gives the researcher an idea of what themes are important and what category names the node tree may contain. A good example of when a researcher may use this method would be in an historical document or newspaper search.

Step 3 - Creating *nodes* and beginning the coding process

In NVivo *nodes* are data categories or files. You can create nodes to reflect the topics of data that you have. Organising data into nodes means you can *store, retrieve and analyse* all the data. Nodes may be expected topics or emergent topics. Many researchers begin by creating nodes that reflect individual interviews . You can code data to the node for that topic by allocating it to an existing node or by creating new node for it.

NVivo also has an *autocoding* function but you will still need to check the autocoding to ensure that each piece of data has been correctly assigned. To autocode information within NVivo, write a query to tell NVivo how to allocate the data. Then run the query and then check that NVivo generated the required information. To do this, in the *Navigation view*

Click the *Queries* folder.

On the *Explore tab*, select the *New Query*.

Select *Text search query*.

The *text search dialog box* opens up.

Tick *add to project*.

When you do this, the *General tab* will become visible.

Give the *query* a name.

On the *Explore tab*, select *Run Query*.

Scroll through the results. You should have coding references followed by coded text.

As Shown in Figure 3 shows the creation of a new node. Each node is assigned a *name*. You can also assign a *descriptor* of the node. This is important where the project is likely to be of some duration or where there is more than one person accessing the project file.

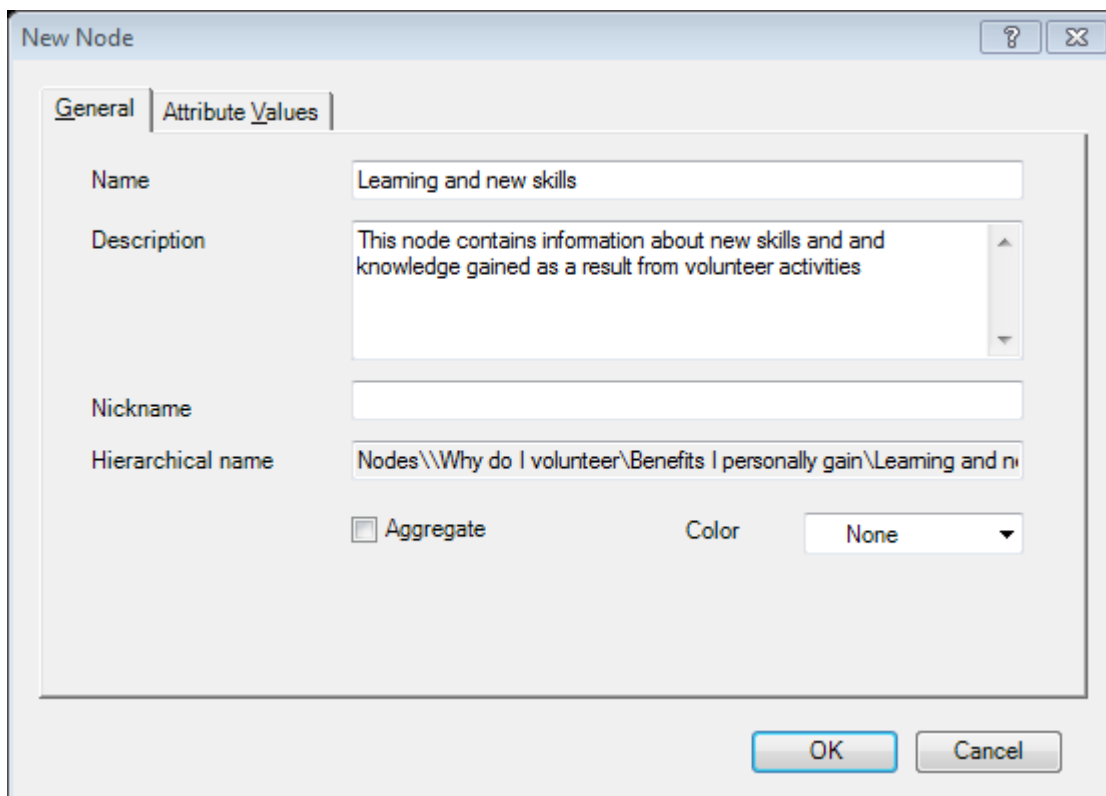


Figure 3 – New node creation with descriptor

You can organise the data into different node arrangements or folders. A *node tree* is a hierarchical structure of major and sub categories or folders. There are also *free nodes*. These are data files for which there is no obvious link, but you know the data may be important at a later date. Free nodes are used for storing *deviant data*.

Deviant data is data that does not appear to fit within the categories being developed.

Many researchers interview on specific themes. As noted in Chapter 14, they have conducted a literature search prior to the qualitative research stage to determine the

themes that participants will be interviewed on. Where this is the case, the researcher may already have an idea of what codes or node names to use. This is called *a priori coding*. Even where *a priori coding* has been undertaken, the researcher may still not know how the key themes relate to each other.

Figure 4 shows a node tree with two parent nodes. The first parent node labelled 'Participant' relates to the original de-identified interview participants. The node contains only the information that relates to each individual person. It contains the transcribed interviews, the original wave files and imported Excel spread-sheet data containing information on age group, current voluntary activity and educational characteristics for some individuals, and employment status. The second parent node labelled 'Why do I volunteer' contains specific thematic passages from the interviews. The nodes seen under each parent node are called *child nodes*. Child nodes contain sub categories within the theme. So the parent node 'participant' contains all the transcripts whereas the child node 'Anne' contains only the interview transcript of Anne. Coding in this way ensures that the original data set stays with the original participant and is uncorrupted. The researcher can always go back to the original full transcripts.

The screenshot displays the NVivo software interface. The main window shows a node tree on the left and a list of nodes on the right. The node tree has two parent nodes: 'Participant' and 'Why do I volunteer'. Under 'Participant', there are child nodes for individual names: Anna, Bernadette, Fredric, Grace, Ken, Mary, Nick, Phoebe, and Sunil. Under 'Why do I volunteer', there are child nodes for thematic categories: 'Benefits I personally gain', 'Centres person', 'I can make a difference', 'I have the time', 'Learning and new skills', 'Recognition', 'Social experience', 'There is personal attachment', 'Benefits others got from my volun', 'Other reasons for volunteering', and 'Word search result cloud'. The list of nodes on the right provides details for each node, including Name, Sources, References, Created On, Created By, Modified On, and Modified By. The interface also includes a menu bar (File, Home, Create, External Data, Analyze, Explore, Layout, View) and a toolbar with various functions like Code Selection At, New Node, Existing Nodes, Range Code, Spread Coding, Uncode Selection At, This Node, Existing Nodes, Uncoding, Memo Link, See Also Link, Hyperlink, New Annotation, Delete Annotation, New Summary Link, Delete Summary Link, Auto Summarize, Framework Matrix, Classify Nodes from Dataset, Classification, and Stop Words. The sidebar on the left contains navigation options: Nodes, Relationships, Node Matrices, Sources, Nodes, Classifications, Collections, Queries, Reports, Models, and Folders. The bottom of the interface shows a search bar and a list of nodes, with the current node being 'Why do I volunteer'.

Figure 4 – The node tree

It is possible to investigate the link between themes within NVivo. In the following screenshot we can see who discussed the concepts of *volunteering* and *recognition* for example. Figure 5 shows the way models are used to indicate pictorially the relationship between an idea, in this case 'recognition' and who actually spoke about recognition.

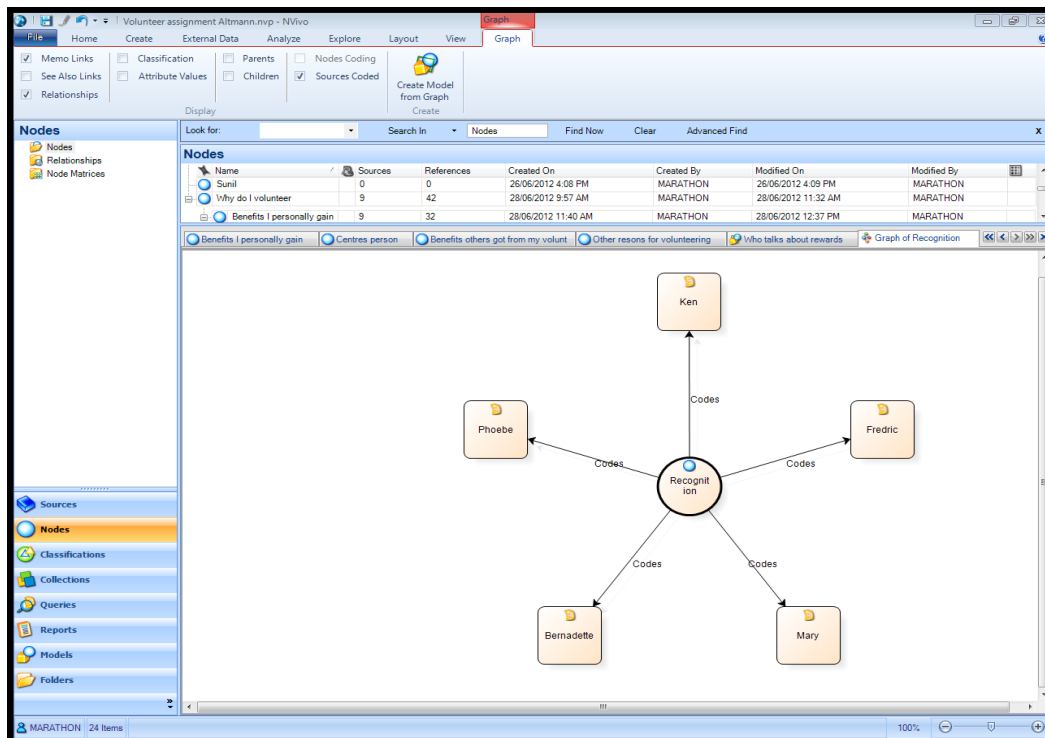


Figure 5 – pictorial model showing who spoke about the concept of recognition.

Thematic linking is one way of considering the relationship between ideas and the possible *nodes or folders* are set up. This is useful when determining the coding node hierarchy. This aligns with Ezzy's (2002, p. 89) view that open coding is exploratory and therefore inductive. However because the interview was structured, Hay (2005) indicates that, it is possible to pre-code the data using a 2 step process beginning with basic coding in order to distinguish overall themes, followed by a more in depth, interpretive code in which more specific trends and patterns can be interpreted. When using theme as a coding unit, Minichiello et al (1990) considers coding to be equivalent to looking for the expressions of an idea in which a text

chunk represents a single theme or issue of relevance to your research question(s). Most qualitative researchers use a combination of *a priori* and *inductive coding*

In order to begin the coding process, a *codebook* needs to be developed. The codebook is a valuable tool in defining each node, what to code to each node, what not to code to a particular node and provides examples appropriately coded material. Figure 5 provides an excerpt from the codebook used in the volunteer research project.

The code	Brief definition	When to use the code	Example of when to use this code	When not to use	Example of when not to use
Why do I volunteer	Defines reasons why each individual volunteers.	When each person identifies a reason to volunteer. Often but not always denoted by the use of I. Each participant may list several different reasons for volunteering. Each reason is coded separately.	I feel like my work has made a difference.	Do not code to this node if the person is talking about general reasons for volunteering, or when discussing other people who have volunteered	It is a way people can share their abilities and skills to benefit the community
Benefits I personally gain from volunteering	Defines personal benefits that each person has received from volunteer activity.	Use this when each person identifies a benefit that they personally receive from volunteering. Each participant may list more than benefit. Code each benefit separately.	I felt I was doing something of use for the church	Do not use when person is not defining a benefit. Where a participant is defining a benefit, it must relate to a benefit the participant personally receives. Do not use when participant is describing a benefit that other receive. When it is a statement about what the participant does.	I also volunteer for half a day a week Or It's a way people can share their abilities and skills

Figure 6 - code book example for volunteer project

Creating structure around the coding process in this way increases validity, reliability and rigorousness of the research project and is in line with Blanchard et al's (1997) guidelines for coding.

In large projects where coding is conducted by more than one person, maintaining *validity* and *reliability* in coding is important. *Validity* measures the relationship between the concept and the code and *reliability* is concerned with accuracy, stability and reproductability. Both increase the rigorousness of the research. By ensuring

concept match to the code, providing definitions and examples of coded information, sources of error were potentially reduced. Reliability of the coding process can be checked through NVivo's *coding comparison query option*. Within Nvivo, the *Kappa value* can measure coding stability, accuracy and inter-coder reliability. The kappa Value is a statistical coefficient measure between two or more sets of data. It is useful where more than one person has been responsible for the coding for example.

Step 4 - Exploring themes within NVivo

In step 2 you considered the coding of transcripts to explore how codes fitted together into categories, that is *factual* and *topical* coding. Factual coding is used to store known facts about data items. These may include participant gender or age for example, or information about the interview setting. Topical coding is used to identify material on a topic for later retrieval and description, categorisation or reflection. Factual and topical coding are used in the early generation of themes and nodes. Whereas analytical coding occurs as a later step. Analytical coding is used to develop themes in theoretical interpretations, for example theorising about existing concepts, identifying new concepts and constructing concepts. It can also be used to theorise about association or relationships between concepts. Some of the codes were generated using theme emergence from the data. The early thematic analysis and coding described a category. It did not provide an explanation or interpretation of the data.

Here in Step 4, consideration is given to identifying particular ideas and concepts that assist in the explanation of the data. These are also called themes and build upon the earlier inductive analysis provided by NVivo and the coding process. Think of it like this. You go to the sports pages of a newspaper. In step two you might thematically sort the news articles into different sports categories (soccer, hockey, netball) for example. In step four, you sort according to sports injury and sports injury type. The step four analysis considers the relationship between sports injury type and the actual sport being played at the time of the injury. An example from the NVivo volunteer project is relevant here.

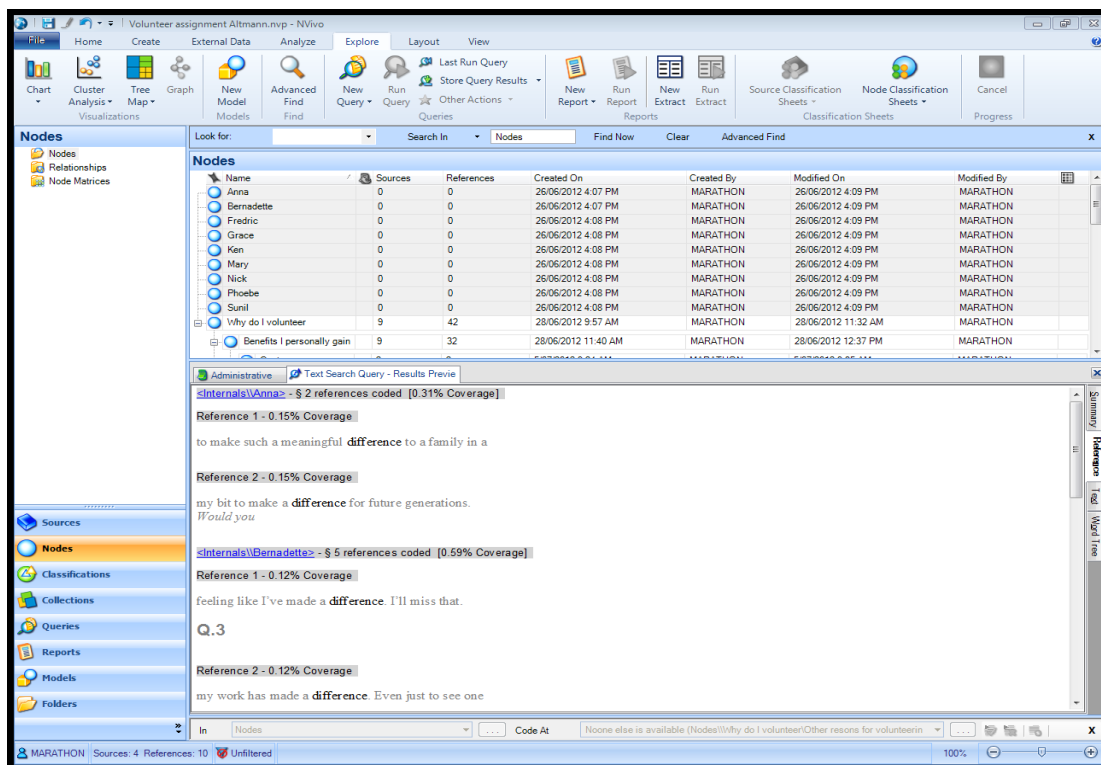


Figure 7 - text query result on the word difference'

As shown in figure 7, a text query on the word 'difference' was run for all the imported participant transcripts in order to determine whether all the interviews containing the word 'difference' had been picked up under the node 'making a difference'. This increased validity, but also allowed the researcher to consider the context in which the word 'difference' was used. It is possible to determine the exact context of the word use by switching between the *summary tab* and the *reference tab*. The word is used in two distinct ways. First it was used in the context of others, volunteering 'made a difference to a family member'. Whereas Bernadette used the word 'difference' in reference to self rather than others.

Step 5 - Develop node system to illustrate their relationships

After coding, the entire data set still needs to be rechecked for consistency. This is to ensure that material from earlier codes is in line with the final node tree since new codes are added throughout the coding process to reflect the analytical and interpretation processes. The new node tree will reflect the additional relationships as either parent or child nodes. While not shown here, relationship models can be

built to provide a visual representation of the inter relationships or dependencies between the data.

Step 6 - Reporting and Write up

While many of the report functions, queries and models can be used to facilitate report writing processes. In larger research projects monthly, bi-monthly, quarterly or annual progress reports are produced. Researchers are able to set up report templates within NVivo to produce and reproduce reports at these intervals.

However the final write up of research is usually undertaken external to NVivo. The report writing process is facilitated by wide range of export formats available but the thematic models produced within NVivo may enhance the production of professional reports.

Strengths and weaknesses of NVivo

Strengths

The software is multi use in that it provides allows both inductive thematic coding and deductive code and retrieve functions. The Document and Node systems mimic the logic and interface of Microsoft Windows Explorer, so that NVivo seems intuitive to use compared to other qualitative packages. One of the greatest strengths of NVivo is its capacity to interrogate large amounts of qualitative data to identify themes and relationships.

Compared to other computer assisted qualitative data analysis packages such as NUD•IST and ATLAS.ti, NVivo allows interface with a wider range of other software packages. This allows better transferability of data for generating final reports. Many of the images and data strings can be imported directly into other software packages such as Microsoft Word or Excel for example. Because of the wide range of import and export opportunities, NVivo provides limited support for researchers working with mixed methods. For example, you can analyse qualitative survey data and then exchange the data with applications like Microsoft Excel and IBM SPSS Statistics, thus enabling comparisons based on demographic data.

The search and query tabs allow the user to search for either specific text passages or identify groups of data. New nodes can then be created to store the information

based on new themes or relationships. It does not corrupt the original source.

Deviant sources of information can be stored for later testing

When all functions are used, NVivo becomes a central place to store all project records. Specific report formats can be created and re-run at intervals to assist in higher level reporting required by all research projects. The large range of import and export options within NVivo facilitate this.

Weakness

There is a danger that qualitative researchers consider the descriptive thematic coding of data to be the end of the project and fail to interpret the data adequately. This is still a 'brain job' that occurs outside of the software program. The researcher remains the main tool for analysis (Denzin & Lincoln 2005). There is danger in overestimating the software capabilities. Computers do not perform independent rational process or substitute the analyst's interpretative capacities.

Where voice or film files are imported into NVivo as base data, time pressures may mean that coding is done from the original voice data and transcripts of the coded data generated. In making the final interpretation of the issue under investigation, a skewed result may be achieved as a result of not linking the data back to the full transcript.

The coding function relies heavily on the creation of a valid, rigorous code book. Where insufficient attention is given to this aspect (which cannot be undertaken in NVivo), the validity, reliability and replicability of the study may be jeopardised.

When generating models as in figure 5, the user needs to be aware that the model is not updated as new information is added or recoded within NVivo. This is not the case with other qualitative software such as Nudist for example.

NVivo also relies on the user maintaining the hierarchical node structure in order to explore relationships under different contexts. Where this is not maintained, there is a tendency for the data relationships to become skewed.

Computer assisted qualitative data analysis software is generally not suited to small amount of data. Analysis of small amounts of qualitative data is more easily

undertaken by an experienced researcher manually. The time for novice users to become familiar with the software will also impact on project timelines.

Summary and exercises

This chapter has considered how computer assisted data analysis is of assistance to qualitative researchers. It did this by focusing on one particular qualitative software package – NVivo. It provided a stepped process that introduced terminology that qualitative researchers need to become familiar with. The strengths and weaknesses of NVivo have been outlined. Many of these relate to all computer assisted qualitative software packages. The following exercises may be helpful to the students and researchers seeking to learn more about the NVivo program.

Exercise 1

Log onto http://www.qsrinternational.com/products_free-trial-software.aspx and view the online promotional video or take a live e demo at http://www.qsrinternational.com/products_free-trial-software.aspx.

Exercise 2

Log onto http://www.qsrinternational.com/products_free-trial-software.aspx and download the 30 day, free trial software. The software allows you to explore features of NVivo using a supplied data set.

References

- Blanchard, L 1997, 'How Do You Define Community? Perspectives of Community Members', Paper presented at the American Anthropological Association Annual Meeting, 19-23 November, Washington, DC.
- Denzin, NK & Lincoln, YS 2005, 'Introduction: The discipline and practice of qualitative research'. In N K Denzin & YS Lincoln (Eds.), *The Sage handbook of qualitative research* (3rd ed. pp.1–32). Thousand Oaks, CA: Sage.
- Ezzy, D 2002, *Qualitative Analysis Practice and Innovation*, Crows Nest, NSW, Allen & Unwin.
- Geertz, C 1973, *Thick Description: Toward an Interpretative Theory of Culture.* In *The Interpretation of Cultures*, New York, New York Basic Books.
- Hay, I 2005, 'Qualitative research methods in human geography', 2nd ed, Oxford University Press, Oxford.
- Minichiello, V, Aroni, R, Timewell, E, & Alexander, L 1990, *In-Depth Interviewing: Researching People*. Hong Kong: Longman Cheshire
- Rademaker, LL, Grace, EJ & Curda, SK 2012, 'Using Computer-assisted Qualitative Data Analysis Software (CAQDAS) to Re-examine Traditionally Analyzed Data: Expanding our Understanding of the Data and of Ourselves as Scholars'. *The Qualitative Report*, Vol 17, Is 43, pp. 1-11. Available at <http://www.nova.edu/ssss/QR/QR17/rademake.pdf>
- Willis, K 2002, 'Analysing Qualitative Data'. In M Walters (Eds), *Social research Methods* (2nd Ed. Pp. 416-421), Oxford University Press, Australia & New Zealand.